Journal of Public Economics 165 (2018) 101–113

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Climate agreements in a mitigation-adaptation game ☆

Basak Bayramoglu[a,*], Michael Finus[b,c], Jean-François Jacques[d,e]

[a] Economie Publique, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France
[b] Department of Economics, University of Graz, Austria
[c] University of Bath, UK
[d] Université Paris-Est, ERUDITE (EA 437), UPEM Marne la Vallée 77454, France
[e] Université Paris-Dauphine, Université PSL, LEDa, [CGEMP], Paris 75016, France

ARTICLE INFO

Article history:
Received 13 October 2016
Received in revised form 22 March 2018
Accepted 10 July 2018
Available online 26 July 2018

JEL classification:
C71
D62
D74
H41
Q54

Keywords:
Climate change
Mitigation-adaptation game
Public good agreements
Strategic substitutes versus complements

ABSTRACT

We analyze the strategic interaction between mitigation (public good) and adaptation (private good) strategies in a climate agreement. We show the fear that adaptation will reduce the incentives to mitigate carbon emissions may be unwarranted. Adaptation can lead to larger self-enforcing agreements, associated with higher global mitigation levels and welfare if it causes mitigation levels between different countries to be no longer strategic substitutes but complements. We argue that our results extend to many public goods. The well-known problem of "easy riding" may turn into "easy matching" if the marginal utility of public good consumption is strongly influenced by private consumption.

1. Introduction

Climate change is probably one of the most important challenges of mankind. The Kyoto Protocol signed in 1997 was the first global treaty with specific mitigation targets but turned out to be not sufficient to address global warming. After several years of negotiations, a successor agreement was recently signed in Paris in 2016. However, most scholars doubt that the Paris Accord will be sufficient to keep the increase of the global surface temperature below 2 °C, a widespread accepted target to avoid severe interference with the climate system.

☆ Some work on this paper was carried out while Michael Finus was visiting the CGEMP Université Paris-Dauphine, Université Paris-Est Marne-la-Vallée, and INRA UMR Economie Publique, and while Basak Bayramoglu and Jean-François Jacques were visiting the University of Bath. Logistic and financial support by these institutions is gratefully acknowledged. Michael Finus benefitted from the EU-FP7 project ECONADAPT, Grant Agreement No: 603906. The authors are also grateful for very helpful comments by the Editor and two anonymous referees. We also thank Françoise Forges for detailed comments. Helpful comments have also been received at the following conferences/workshops/seminars: Workshop on Climate Change and Public Goods, Fondazione Eni Enrico Mattei, Venice (June 2014), the 5th World Congress of the Association of Environmental and Resource Economists, Istanbul (June–July 2014), workshop RECAP 15, "Step-by-Step — How to Progress in International Climate Change", Berlin (February 2015), The Second Environmental Protection and Sustainability Forum, University of Bath, Bath (April 2015), the 2nd FAERE Conference, Toulouse (September 2015), the Research Workshop on Contemporary Topics in Economics of the Environment and Sustainability, NTU, Singapore (March 2016), the 23rd Annual Conference of the EAERE, Athens (June–July 2017), the 15th Annual Congress of the EAAE, Parma (August 2017), the 67th AFSE-Meeting, Paris (September 2018), the Economic Seminar, Chaire Economie du Climat, Paris (March 2015), the Economic Seminar, PEEES, Paris (June 2015), the Economic Seminar at the University of Wageningen, Wageningen (February 2016), the Economic Seminar, University of Kent, Canterbury (October 2016), and the Economic Seminar, CIRED, Nogent-sur-Marne (February 2016).

* Corresponding author.
E-mail addresses: Basak.Bayramoglu@inra.fr (B. Bayramoglu), michael.finus@uni-graz.at (M. Finus), Jean-Francois.Jacques@u-pem.fr (J.-F. Jacques).

https://doi.org/10.1016/j.jpubeco.2018.07.005

Clearly, mitigation to address the cause of global warming is costly, participation in a climate treaty is voluntary and compliance is difficult to enforce. Due to the slow progress of curbing global warming, and the first visible impacts of climate change, in particular in developing countries, adaptation measures (like building dykes against flooding and installing air-conditioning devices against heat) have received more attention in recent years. This is reflected in the negotiations leading to the Paris Accord but also in the scientific community, as for instance summarized by various recent reports by the Internal Panel on Climate Change (IPCC). In contrast to mitigation (i.e. reduction of emissions), which can be viewed as a non-excludable public good, adaptation (i.e. amelioration of climate damages) is usually viewed as a private good; it only benefits the country in which adaptation measures are implemented. The key research question which we try to answer is: how does adaptation, as an additional strategy to mitigation, affect the prospects of international policy coordination to tackle climate change?

At the outset, the answer is not straightforward when considering the following points. The "pessimists" argue that adaptation will shift the focus away from mitigation. Since typically the benefits from mitigation are lower in the presence of adaptation, equilibrium mitigation levels will be lower. Thus, also the positive welfare externalities across countries from cooperation are lower. The "optimists" point out that lower mitigation levels reduce the incentive to free-ride and hence larger agreements may be stable. Moreover, having a second strategy reduces the costs of addressing a global externality.

We show that the arguments of optimists and pessimists are correct, on balance, optimistic factors dominate the outcome, but the main driver for optimism is a very different one. In the presence of adaptation, reaction functions in mitigation space may be upward sloping. That is, mitigation levels in different countries may no longer be strategic substitutes but complements. Such a matching behavior makes it easier to form large stable coalitions in order to increase public good provision, which in most cases leads to larger global welfare.

The possibility of upward-sloping reaction functions arises when the substitutional or complementary relationship between equilibrium mitigation and adaptation in a country is sufficiently strong (to which we refer later as cross or indirect effect) compared to the substitutional relationship between own and foreign mitigation levels (to which we refer later as own or direct effect). It is interesting that this possibility does neither need the assumption that mitigation and adaptation are substitutes as is commonly believed (they can also be complements) nor does a large indirect effect need to imply that the sufficient conditions for the existence and uniqueness of equilibria are violated.

Our paper is related to four strands of literature. Firstly, there is large body of literature on the game-theoretic analysis of international environmental agreements (IEAs), which can be traced back to Barrett (1994) and Carraro and Siniscalco (1993), with recent publications for instance by Harstad (2012), El-Sayed and Rubio (2014) and Battaglini and Harstad (2016).[1] Only some recent papers have studied mitigation and adaptation in a strategic context. Different from for instance Buob and Stephan (2011), Ebert and Welsch (2011, 2012), Zehaie (2009) and Eisenack and Kähler (2016), we allow for more than two players and study the formation of agreements. Different from some recent work by Barrett (2008) and Benchekroun et al. (2017) who study IEAs, we work in a much more general framework, allow for the possibility that mitigation and adaptation could not only be substitutes but also complements and derive most results analytically. Also different from them, we allow for the possibility of strategic complementarities in mitigation space, which is an important factor for generating interesting results.

Secondly, there is a literature on non-convexities of negative externalities, including early contributions by Baumol and Bradford (1972), Laffont (1976) and Starrett (1972). This literature does not consider agreement formation but points to the fact that private actions can ameliorate the damage from negative externalities. Noticing that any public bad game can be recasted in a public good game framework, where the latter is the setting of this paper, this means non-concavity of positive externalities. We show that in the presence of the amelioration of climate damages through adaptation, the conditions for upward-sloping reaction functions in public good provision space are exactly those related to the non-concavity of an agent's payoff function with respect to other players' provision levels, in line with this strand of literature.

Thirdly, there is a large literature on the private provision of (pure) public goods (e.g. Bergstrom et al., 1986; Cornes and Hartley, 2007; Fraser, 1992). "Private" means non-cooperative with the possibility of cooperative agreements normally not being considered in this literature. Typically, agents maximize a utility function subject to a linear budget constraint, with utility being derived from the total level of public provision (which is the sum of individual contributions) and a private numeraire good. The standard assumption is that both goods are normal goods, and the cross derivative of utility with respect to the public and private good is assumed to be of minor importance. This gives rise to downward sloping reaction functions in public good provision space and a unique equilibrium public good provision vector. However, downward sloping reaction functions, usually associated with the term "easy riding" (Cornes and Sandler, 1984), is not the only possibility as pointed out by Cornes and Sandler (1986, ch. 5). Moreover, it does not seem unrealistic to consider the possibility that some public goods are superior goods, like some environmental goods for which the income elasticities has been reported to be larger than 1. Our model essentially captures this possibility. Our utility function is called a benefit function but is essentially the same. What is different is that we do not assume a linear budget constraint with constant prices, but, in the tradition of the IEA-literature, consider the more general case of (strictly) convex cost functions of private and public good provision and hence non-constant marginal costs.[2] We have downward or upward sloping reaction functions, depending on the relative strength of own and cross effects on benefits, though importantly, only the absolute value (and not the sign) of the cross effect matters for the slope.

Fourthly, there is quite some literature that investigates complementarities in strategic games. From the survey by Vives (2005), it appears that complementarity does not need to be the result of special assumptions but there are many interesting economic problems with this feature, though the analysis is usually more complex, requires different tools for the analysis and may suffer from multiple equilibria. For our problem, it turns out that a slight modification of standard theorems is sufficient for the analysis and a simple condition gives existence and uniqueness of equilibria in both stages of the game.

It is important to note that the possibility of large cross effects on the benefits of public good and private good provision extends much beyond the context of this paper. For instance, member states of the European Community can either coordinate on policy issues like security, anti-terrorism, migration and social policy or pursue those issues nationally. That is, financial resources can either be transferred to Brussels (which can be interpreted as public good provision) or remain with national governments (which can be viewed as private good provision). In practice, national and international

[1] See Finus and Caparrós (2015) for a collection of some of the most influential papers in the field, including a comprehensive overview.

[2] This generalization comes at the cost that the problem can no longer be viewed in terms of income elasticities.

policy measures co-exist and the benefit of national (international) policy measures is often diminished by the quality of international (national) measures.

But also at the national or regional level, citizens can vote for an increase of the local policy force or invest in devices to secure their private homes against burglaries. Farmers can invest either in their own machinery and irrigation devices or in the membership of a cooperative with access to shared facilities. In all these examples, it is likely that the benefit of the private investment impinges on the benefits of the public investment and vice versa, i.e. the cross effect is negative. In other cases, it can be expected that the cross effect is positive. Public spending on improved infrastructure may increase the value of houses and hence makes the private investment in flood protection and security more valuable to home owners.

In what follows, we set out our model and its assumptions in Section 2. We present results of our two stage coalition formation model in reverse order according to backwards induction in Sections 3 and 4, respectively. We evaluate the overall equilibrium in Section 5. Section 6 summarizes our main results and policy conclusions.

2. Model

We consider n players, which are countries in our context, $i = 1, 2, \ldots, n$, with the payoff function of country i in the pure mitigation game (M-game) given by:

$$\Pi_i(Q, q_i) = B_i(Q) - C_i(q_i) \tag{1}$$

and in the mitigation-adaptation game (M + A-game) by:

$$\Pi_i(Q, q_i, x_i) = B_i(Q, x_i) - C_i(q_i) - D_i(x_i). \tag{2}$$

We denote the set of players by N. In the richer M + A-game, country i can not only choose its individual mitigation level q_i but also its adaptation level x_i within its (compact and convex) strategy space $q_i \in [0, \bar{q}_i]$ and $x_i \in [0, \bar{x}_i]$ with $\bar{q}_i$ and $\bar{x}_i$ sufficiently large. Country i's payoff comprises benefits, B_i, which depend on total mitigation, $Q = \sum_{j=1}^{n} q_j$, and in the M + A-game additionally also on its individual adaptation level, x_i; the cost of mitigation is denoted by C_i, and the cost of adaptation by D_i.

Comparing Eqs. (1) and (2), it is evident that the M + A-game can be viewed as a generalization of the M-game. A direct conversion for comparison of welfare is possible when normalizing $B_i(Q, x_i)$ in the M + A-game such that for zero adaptation, $x_i = 0$, benefits in both games are equal, $B_i(Q, 0) = B_i(Q)$ and by setting adaptation costs to zero for $x_i = 0$, $D_i(0) = 0$. Both assumptions do not appear to be strong assumptions. In the following analysis, we do not require these assumptions, except if we compare payoffs in the two games in Section 5.

If there is no misunderstanding, we drop the index i as we assume that players are ex-ante symmetric, i.e. they have the same payoff function; if we need to stress that players are ex-post asymmetric, e.g. because they chose different strategies, we will use the index. As will become apparent below, signatories and non-signatories will typically choose different mitigation levels and hence will receive different payoffs. The assumption of ex-ante symmetric players is very much in the tradition of the literature on coalition formation in general (see Bloch, 2003; Yi, 1997 for overviews) and on IEAs in particular (see Finus and Caparrós, 2015 for an overview) due to the complexity of the analysis of coalition formation.

We assume that all functions, including their first and second derivatives, are continuous in their variable(s). Moreover, we make the following assumptions regarding the components of the payoff functions (with the understanding that all derivatives with respect to x are only relevant in the M + A-game) where subscripts denote derivatives, e.g. $B_Q = \frac{\partial B}{\partial Q}$ and $B_{QQ} = \frac{\partial^2 B}{\partial Q^2}$ and where we drop the arguments in these functions if no misunderstanding is possible.

General Assumptions

Both Games *:*

a) $B_Q > 0$, $B_{QQ} \leq 0$, $C_q > 0$, $C_{qq} > 0$.

b) $\lim_{Q \to 0} B_Q > \lim_{q \to 0} C_q > 0$.

M + A-Game:

c) $B_x > 0$, $B_{xx} \leq 0$, $D_x > 0$, $D_{xx} \geq 0$.
If $B_{xx} = 0$ *, then* $D_{xx} > 0$ *and vice versa: if* $D_{xx} = 0$ *, then* $B_{xx} < 0$.

d) $\lim_{x \to 0} B_x > \lim_{x \to 0} D_x > 0$.

e) *i)* $B_{xQ} = B_{Qx} < 0$ *or ii)* $B_{xQ} = B_{Qx} > 0$.

From a technical point of view, assumptions *a* and *c* reflect the standard assumptions of concave benefit and convex cost functions. We allow for the possibility that benefit functions can be linear such that we can revisit some payoff functions, which have been considered in the literature on IEAs in the context of a pure mitigation game. We assume cost functions of mitigation to be strictly convex in order to ensure unique equilibrium mitigation levels. For adaptation, it turns out that this is not necessary. However, in assumption *c*, we state that if benefit functions are linear in adaptation, then adaptation cost functions must be strictly convex and vice versa. These properties of the benefit and cost functions together with assumptions *b* and *d* rule out corner solutions as for instance in Kolstad (2007) in a pure mitigation game and in Barrett (2008) in a mitigation-adaptation game.

From an economic point of view, assumption *a* stresses that mitigation is a pure public good, i.e. the marginal benefit from mitigation depends on the sum of all (and not on individual) mitigation efforts. In contrast, assumption *c* stresses that adaptation is a private good, i.e. the marginal benefit from adaptation depends on the individual adaptation level of a country (and not on those of others). The interdependency between mitigation and adaptation is captured through assumption *e*. As pointed out in the Introduction, in line with all the literature we are aware of, we consider assumption *ei*) as the standard assumption: the marginal benefit from mitigation (adaptation) decreases with the level of adaptation (mitigation). For completeness, and to stress the robustness of our results, we consider also assumption *eii*), which could be relevant in the context of other economic problems of public-private good provision.[3] For simplicity, the interdependency between mitigation and adaptation is assumed away on the cost side. In order to stress this, we assume for clarity two separate cost functions.

The strategic interaction between countries is directly related to the (pure) public good nature of mitigation. Mitigation in country i generates benefits in country i but also in all other countries. Thus, mitigation levels generate positive externalities. Adaptation levels generate no direct externalities. However, they indirectly influence the strategic interaction among countries. If the cross derivative is negative (positive), $B_{Qx} < 0$ ($B_{Qx} > 0$), mitigation and adaptation are substitutes (complements): the higher the adaptation level in a country, the lower (higher) will be its mitigation level, irrespective whether country i acts independently or joins an agreement.

We assume the General Assumptions to hold throughout the paper. If we make further assumptions, we will mention them explicitly. Our two-stage coalition formation game unfolds as follows.

[3] Without loss of generality, we avoid the economically uninteresting case of $B_{Qx} = 0$ to simplify the subsequent exposition.

Definition 1 (Coalition formation game).

Stage 1. All countries choose simultaneously whether to join coalition $P \subseteq N$ or to remain a singleton player. Countries $i \in P$ are called signatories and countries $j \notin P$ are called non-signatories.

Stage 2. All non-signatories $j \notin P$ choose their economic strategies in order to maximize their individual payoff and all signatories $i \in P$ do so in order to maximize the aggregate payoff to all coalition members. Choices of all players are simultaneous.

Stage 1 is the cartel formation game, which originates from the literature in industrial organization (d'Aspremont et al., 1983) and has been widely applied in this literature (e.g. Deneckere and Davidson, 1985; Donsimoni et al., 1986; Poyago-Theotoky, 1995; see Bloch, 2003; Yi, 1997 for surveys) but also in the literature on IEAs (e.g. Barrett, 1994; Carraro and Siniscalco, 1993; Rubio and Ulph, 2006; see Finus and Caparrós, 2015 for a survey). This game has also been called open membership single coalition game as membership in coalition P is open to all players and players have only the choice between joining coalition P or remaining a singleton. Open membership may be defended on two grounds. In the context of the provision of public goods, it appears that one is more concerned about players leaving a coalition than joining it. Moreover, to the best of our knowledge, all international environmental treaties are of the open membership type. The assumption of a single coalition simplifies the analysis but is also in line with the historical records of IEAs with a single treaty.

Stage 2 follows the standard assumption in the literature on coalition formation (see Bloch, 2003; Yi, 1997 for surveys): the coalition acts as a kind of meta-player (Haeringer, 2004), internalizing the externality among its members, whereas non-signatories act selfishly, maximizing their own payoff. We also follow the mainstream assumption and assume that signatories and non-signatories choose their economic strategies simultaneously. In the M-game, the second stage delivers an equilibrium mitigation vector $\boldsymbol{q}^*(P)$, given that coalition P has formed. In the M + A-game, $\boldsymbol{q}^*(P)$ and $\boldsymbol{x}^*(P)$ are derived.[4]

The two-stage coalition formation game is solved by backwards induction. It is clear that we want for practical reasons for each possible coalition P, a unique equilibrium strategy vector to exist. This allows us to write $\Pi_i^*(P)$ instead of $\Pi_i(\boldsymbol{q}^*(P))$ in the M-game, and, accordingly, $\Pi_i^*(P)$ instead of $\Pi_i(\boldsymbol{q}^*(P), \boldsymbol{x}^*(P))$ in the M + A-game. Even though we provide sufficient conditions for existence and uniqueness only in the next section, we make already use of this assumption in order to save on notation and define a stable coalition P^* as follows:

$$\begin{aligned} &\text{internal stability:} \quad \Pi_i^*(P^*) \geq \Pi_i^*(P^* \backslash \{i\}) \quad \forall i \in P^* \\ &\text{external stability:} \quad \Pi_j^*(P^*) \geq \Pi_j^*(P^* \cup \{j\}) \quad \forall j \notin P^*. \end{aligned}$$

It is evident that the conditions of internal and external stability de facto define a Nash equilibrium in membership strategies in the first stage. Each player i who announced to join coalition P^* should have no incentive to (unilaterally) change her strategy by leaving coalition P^* and each player j who announced not to join coalition P^* should have no incentive to (unilaterally) change his strategy and join coalition P^*, given the equilibrium announcements of all other players.

[4] Note that all our results also hold if mitigation is chosen first and then adaptation, as proved in our working paper version. This would be different if adaptation is chosen first and then mitigation, as for instance considered in Zehaie (2009). For a discussion, see the concluding Section 6.

Note that by construction, the equilibrium economic strategy vectors in the second stage correspond to the Nash equilibrium known from games without coalition formation if coalition P is empty or contains only one player. We also call this "no cooperation" . By the same token, if coalition P comprises all players, i.e. the grand coalition forms, $P = N$, this corresponds to the "social optimum" . We also call this "full cooperation" . Any non-trivial coalition (i.e. a coalition of at least two players) which comprises more than one player but less than all players may be viewed as partial cooperation.

3. Second stage of coalition formation

3.1. Preliminaries

Suppose some coalition $P \subseteq N$ of size p has formed in the first stage. Then the first order conditions in terms of mitigation are given by

$$pB_Q(Q) = C_q(q) \tag{3}$$

in the M-Game and by

$$pB_Q(Q, x) = C_q(q) \tag{4}$$

in the M + A-game with $p = 1$ for non-signatories and $p \geq 2$ for signatories if a non-trivial coalition forms. Consequently, signatories internalize the externality among their group and hence will mitigate more than non-signatories, i.e. $q^*_{j \notin P}(p) < q^*_{i \in P}(p)$ for all p , $1 < p < n$. This holds in both games. In the M + A-game, we have additional first order conditions for adaptation, which are the same for non-signatories and signatories:

$$B_x(Q, x) = D_x(x). \tag{5}$$

Hence, the adaptation level $x^*(p)$ of signatories and non-signatories is the same because adaptation is a private good. However, one should therefore not mistakenly conclude that policy coordination is not required in terms of adaptation. Adaptation influences equilibrium mitigation levels and if both are strategic substitutes, there is too much adaptation (and too little mitigation) from a global perspective as long as the grand coalition does not form.

Now if we let $Q = q_i + Q_{-i}$, each first order condition Eq. (3) in the M-game implicitly defines q_i as a function of Q_{-i}, which is the reaction function of player i. In the M+A-game, the same is true with respect to Eq. (4). We can express everything in terms of mitigation by noticing that the first order conditions (Eq. (5)) implicitly determine adaptation x as a function of total mitigation Q and so we can write $x(Q)$ in Eq. (4). Hence, in both games, we have the best response $q_{i \in P} = r_{i \in P}(Q_{-i})$ for signatories and the best response $q_{j \notin P} = r_{j \notin P}(Q_{-j})$ for non-signatories (setting $p = 1$ in the first order conditions of non-signatories). Moreover, one can view the coalition as one player and because of symmetry all non-signatories as another player. Hence, we can define the aggregate reaction function of signatories by $Q_{i \in P} = r^S(Q_{j \notin P})$ and of non-signatories by $Q_{j \notin P} = r^{NS}(Q_{i \in P})$, with $Q_{i \in P} = pq_{i \in P}$ and $Q_{j \notin P} = [n - p]q_{j \notin P}$ the total mitigation of signatories and non-signatories, respectively, in order to capture the strategic interaction between these two groups in a compact way.

3.2. Results

Our first result establishes existence of a unique equilibrium in the second stage.

Proposition 1 (Existence of a unique interior equilibrium in the second stage). *Consider an arbitrary coalition of size p,* $1 \le p \le n$. *Let* $A^M := B_{QQ}$ *in the M-game and* $A^{M+A} := B_{QQ} + \frac{(B_{Qx})^2}{D_{xx}-B_{xx}}$ *in the M + A-game.*

If $A\left[\frac{p^2}{Cqq(q_{i\in P})} + \frac{(n-p)}{Cqq(q_{j\notin P})}\right] < 1$ *for all players* $i \in N, x_i \in [0, \bar{x}_i]$ *and* $q_i \in [0, \bar{q}_i]$ *hold, then a unique interior equilibrium exists regardless of the sign of A.*

In the M-game $A \le 0$; *in the M + A-game,* $A \le 0$ *or* $A > 0$.

Proof. See Appendix A.1. ■

As it will be apparent shortly from Proposition 2 below, term A determines the value and the sign of the slope of reaction functions. From Appendix A.1, it is apparent that the sufficient conditions are derived based on the concept of replacement functions, which are slightly different than reaction functions, but also their slope depends on term A. In terms of existence and uniqueness, we only need an additional assumption if $A > 0$, which can only happen in the M + A-game, namely, if the cross or indirect effect (i.e. $\frac{(B_{Qx})^2}{D_{xx}-B_{xx}}$) is sufficiently strong compared to the own or direct effect (i.e. B_{QQ}). It is of utmost importance for the following analysis to notice that the sign of A^{M+A} does not depend on the sign of the cross derivative B_{Qx}, but only on its absolute value. It is also important to note that even if the cross effect is stronger than the own effect, this does not necessarily upset the sufficient conditions for uniqueness. Proposition 1 only places an upper bound on the cross effect such that the model is still well-behaved.

Proposition 2 (Slopes of reaction functions in mitigation and adaptation space). *Consider an arbitrary coalition of size p,* $1 \le p \le n$, *and let primes denote the slopes of reaction functions. Further let* $A^M := B_{QQ}$ *in the M-game and* $A^{M+A} := B_{QQ} + \frac{(B_{Qx})^2}{D_{xx}-B_{xx}}$ *in the M + A-game.*

The slopes of individual and aggregate reaction functions of signatories are given by $r'_{i\in P}(Q_{-i}) = \frac{pA}{C_{qq}(q_{i\in P})-pA}$ *and* $r^{S'}(Q_{j\notin P}) = \frac{p^2A}{C_{qq}(q_{i\in P})-p^2A}$, *respectively, and the slopes of non-signatories' reaction functions are given by* $r'_{j\notin P}(Q_{-j}) = \frac{A}{C_{qq}(q_{j\notin P})-A}$ *and* $r^{NS\prime}(Q_{i\in P}) = \frac{(n-p)A}{C_{qq}(q_{j\notin P})-(n-p)A}$, *respectively.*

That is, reaction functions in mitigation space are always weakly downward sloping in the M-game. In the M + A-game, reaction functions are (weakly) downward sloping if $A \le 0$ *and are (strictly) upward sloping if* $A > 0$.

For each possible coalition, the slope of the individual reaction function in mitigation-adaptation space $x = f_{i\in N}(Q)$ *is given by* $f'_{i\in N}(Q) = \frac{B_{Qx}}{D_{xx}-B_{xx}}$ *and hence is downward sloping if* $B_{Qx} < 0$ *and upward sloping if* $B_{Qx} > 0$.

Proof. The derivation follows the same lines as described for replacement functions in the proof of Proposition 1 in Appendix A.1 and is therefore omitted. ■

The first statement sheds light on whether mitigation levels are strategic substitutes or complements. In the M-game, they are always substitutes if we exclude the case $B_{QQ} = 0$ in which case reaction functions are orthogonal. In the M + A-game, this is also the case provided the term A^{M+A} is negative, again with orthogonal reaction functions for the special case of $A^{M+A} = 0$. However, if $A^{M+A} > 0$, then reaction functions are upward sloping and mitigation strategies are strategic complements.[5] Because $B_{QQ} \le 0$, $A^{M+A} > 0$ if $\frac{(B_{Qx})^2}{D_{xx}-B_{xx}}$ is sufficiently large, which captures the interaction between mitigation and adaptation.

Intuitively, this can be understood when considering the first order condition for mitigation $pB_Q(q_i + Q_{-i}, x(q_i + Q_{-i})) = C_q(q_i)$, using $Q = q_i + Q_{-i}$ and the relation between adaptation and mitigation $x(Q)$. Increasing Q_{-i} in a comparative static way (and hence Q) has a direct (or own) negative effect on B_Q, namely reducing B_Q because of $B_{QQ} < 0$. Anything else being equal, this would call for a lower $C_q(q_i)$ in order for the equality to continue to hold and hence a lower q_i because $C_{qq} > 0$. In the M-game, this is the only effect. However, in the M + A-game there is also the indirect (or cross) effect, which increases B_Q and hence calls for a higher q_i. Increasing Q_{-i} increases Q and calls for a lower (higher) $x(Q)$ if $B_{Qx} < 0$ ($B_{Qx} > 0$), which, in turn, increases B_Q. For the inequality to hold, we need a higher $C_q(q_i)$ and hence higher q_i. This second indirect effect is exactly $\frac{(B_{Qx})^2}{D_{xx}-B_{xx}}$. It is important to stress once more that only the magnitude but not the sign of the cross derivative B_{Qx} matters whether reaction functions are upward sloping.

An alternative way of viewing this problem is by noticing that the sign of the second derivative of payoff function Eq. (2) with respect to other players' mitigation levels, after inserting $x(Q)$, depends on the sign of term A. Thus, if $A > 0$, the payoff function is not concave but convex in other players mitigation levels.

Upward sloping reaction functions could lead to more optimistic outcomes in a coalition formation game (i.e. larger coalitions). The intuition is that if mitigation levels are strategic substitutes, any additional increase of signatories' mitigation efforts is countervailed by a decrease of non-signatories' mitigation efforts. In the context of climate change, this has been called (carbon) leakage which makes it less attractive to join an agreement. Thus, upward sloping reaction functions may be viewed as a form of anti-leakage or matching, which may be conducive to form large stable coalitions.

The second statement in Proposition 2 gives a clear answer to the question whether adaptation and total mitigation are substitutes or complements. They are always substitutes, irrespective of the degree of cooperation if the cross derivative B_{Qx} is negative and they are complements if the cross derivative is positive. As argued above, in the context of adaptation the cross derivative is most likely negative, though in the context of other interesting economic problems also the reverse seems possible.

4. First stage of coalition formation

In this section, we analyze stable coalitions. In a first step, we discuss two general properties, which help to explain the difficulties of forming large stable coalitions in both games. It will turn out that there are differences in the two games, but they are not sufficiently pronounced to draw general conclusions about the size and the success of stable coalitions in the two games. Therefore, in a second step, we look at two specific payoff functions, which reveal interesting differences in both games.

4.1. General results

There are two factors which influence the stability of coalitions: a) the incentive to join a coalition, related to the property of superadditivity, and b) the incentive to stay outside the coalition,

[5] The signs of the slopes of reaction and replacement functions are the same as is apparent from Appendix A.1), and only depend on the sign of the term A . The possibility of upward sloping reaction functions has been pointed out by Ebert and Welsch (2011, 2012) in a two-player model.

related to the property of positive (negative) externality. These properties are defined as follows where for notational simplicity we make use of the fact of symmetric players.

Definition 2 (Superadditivity, positive and negative externality).

Let $n \geq p \geq 2$.

(i) The expansion of coalition $p - 1$ to p *is superadditive if:*

$$p\Pi^*_{i\in P\cup\{j\}}(p) > [p-1]\,\Pi^*_{i\in P}(p-1) + \Pi^*_{j\notin P}(p-1)$$

If this holds for all p, $2 \leq p \leq n$ *, the game is a superadditive game.*
(ii) The expansion of coalition $p - 1$ *to* p *exhibits a positive (negative) externality if for all* $j \notin P$ *:*

$$\Pi^*_{j\notin P}(p) > (<)\Pi^*_{j\notin P}(p-1)$$

If this holds for all p, $2 \leq p \leq n$, *the game is a positive (negative) externality game.*

The importance of superadditivity for stability emerges from two facts which are proved in Appendix A.2. Firstly, a necessary (though not sufficient) condition for a coalition of size p to be internally stable is that the move from $p - 1$ to p is superadditive. Secondly, a sufficient condition for the existence of a non-trivial stable coalition is that such a move is at least superadditive when moving from a singleton coalition to a two-player coalition (i.e. according the above notation $p = 2$ and hence $p - 1 = 1$). Clearly, both conditions hold in a superadditive game.

The reason why superadditivity could fail (or its effect is small) can be related to the leakage effect. If the slopes of the reaction functions in mitigation space are negative and steep (term A is negative and the absolute value of A is large) and coalitions are small (and hence there are many outsiders), superadditivity could be violated. In contrast, if reaction functions are upward sloping (A is positive), we have anti-leakage, which is a sufficient condition for superadditivity to hold (see Proposition 3, part *a*, below).

We recall that in the M-game $A^M \leq 0$ and in the M + A-game this may also be true ($A^{M+A} \leq 0$) but also $A^{M+A} > 0$ is possible. Thus, the prospects of existence of a stable and possibly large stable coalition are brighter in the M + A- than in the M-game, even though we need to be aware that superadditivity is only a necessary condition for stability and only one factor which determines stable coalitions.

Proposition 3 (Incentives of cooperation).

a) In both games (M- and M + A-games), a sufficient condition for a superadditive game is $A \geq 0$ *.*
b) The M- and M + A-games are positive externality games.

Proof. See Appendix A.2. ■

A second factor which determines stables coalitions is the nature of the externality which can be illustrated with two observations.[6] Firstly, in superadditive negative externality games, the grand coalition is the unique stable coalition. Both incentives work in the same direction of cooperation. It is beneficial to join an agreement and outsiders are negatively affected by an expansion and hence do not want to stay outside. So all players have an incentive to join an agreement as long as the grand coalition has not been formed.[7] An example are trade agreements which abolish tariffs for members but impose tariffs on imports from non-members.

In contrast, in positive externality games, incentives work in opposite directions and hence precise general predictions are not possible. However, it is well-known from the literature that large coalitions including the grand coalition may not be stable. Examples include output and price cartels and the pure mitigation game. The positive externality is a non-excludable benefit generated by the coalition. Those benefits stem from supporting high prices in cartels and from the pure public good nature of mitigation in the M-game. However, also the M + A-game exhibits positive externalities. Adaptation may weaken the positive externality effect, but it does not change the nature of the externality (Proposition 3, part *b*).

4.2. Specific payoff functions

We consider two specific payoff functions which have been frequently applied in the IEA-literature; we call them payoff functions 1 and 2. Both functions allow for a direct conversion of the M + A-game to the M-game by setting adaptation to zero, $x_i = 0$, as discussed in Section 2. Both payoff functions assume quadratic cost functions. Payoff function 1 assumes a linear benefit function:

$$\Pi^M_{i(1)} = bQ - \frac{c}{2}q_i^2 \tag{6}$$

in the M-game and

$$\Pi^{M+A}_{i(1)} = b(1-\gamma x_i)Q + a(1-\lambda Q)x_i - \frac{c}{2}q_i^2 - \frac{d}{2}x_i^2 \tag{7a}$$

$$\Pi^{M+A}_{i(1)} = b(1+\gamma x_i)Q + a(1+\lambda Q)x_i - \frac{c}{2}q_i^2 - \frac{d}{2}x_i^2 \tag{7b}$$

in the M + A-game where the parameters a, b, c, d, γ and λ are assumed to be strictly positive. For payoff function 1, $A^M = 0$ and $A^{M+A} > 0$. We allow for both possibilities, $B_{Qx} < 0$ (Eq. (7a)) and $B_{Qx} > 0$ (Eq. (7b)). That is, we consider that mitigation and adaptation are substitutes but also that they could be complements.

Payoff function 2 assumes again a linear benefit function in terms of adaptation but a quadratic benefit function in terms of mitigation, such that we have $A^M < 0$ and $A^{M+A} ><= 0$ where the sign of A^{M+A} depends on the parameter values.

$$\Pi^M_{i(2)} = \left(bQ - \frac{g}{2}Q^2\right) - \frac{c}{2}q_i^2 \tag{8}$$

$$\Pi^{M+A}_{i(2)} = \left(bQ - \frac{g}{2}Q^2\right) + x_i(a - fQ) - \frac{c}{2}q_i^2 - \frac{d}{2}x_i^2 \tag{9a}$$

$$\Pi^{M+A}_{i(2)} = \left(bQ - \frac{g}{2}Q^2\right) + x_i(a + fQ) - \frac{c}{2}q_i^2 - \frac{d}{2}x_i^2. \tag{9b}$$

Again, we assume all parameters a, b, c, d, f and g to be strictly positive and, again, we capture $B_{Qx} < 0$ in Eq. (9a) and $B_{Qx} > 0$ in Eq. (9b).

For both payoff functions, we need to impose conditions such that the examples are in line with the General Assumptions and the additional assumption for uniqueness of second stage equilibria as specified in Proposition 1 in the M + A-game, provided $A^{M+A} > 0$. This includes conditions to ensure an interior second stage equilibria for every p, $1 \leq p \leq n$. Those conditions as well as all subsequent results are spelled out in detail in Appendix A.3. Moreover,

[6] For overviews of positive and negative externality games, see Bloch (2003) and Yi (1997).

[7] See Weikard (2009) for a proof.

in case of multiple stable coalitions, we invoke for simplicity the Pareto-dominance criterion, which gives us always a unique first stage equilibrium.

Proposition 4 (Stable coalitions for payoff functions 1 and 2**).**

a) For payoff function 1, $A = 0$ and $p^* = 3$ in the M-game whereas in the M + A-game, $A > 0$ and $p^* \geq 3$ which includes the grand coalition for some values of A.
b) For payoff function 2, $A < 0$ and $p^* = 1$ or $p^* = 2$ in the M-game. In the M + A-game, $p^* \geq 3$ if $A \geq 0$ which includes the grand coalition for some values of A, and $p^* = 1$ or $p^* = 2$ if $A < 0$.

Proof. See Appendix A.3. ■

Proposition 4 stresses that the sign of B_{Qx} does not matter for the size of stable coalitions. Both payoff functions confirm the intuition that if reaction functions are upward sloping in the M + A-game, stable coalitions will be (weakly) larger in the M + A-game than in the M-game. For both payoff functions, $A \geq 0$ implies $p^* \geq 3$ and $A < 0$ implies $p^* < 3$. We recall that the sign of term A does not depend on the sign of the cross derivative B_{Qx} but only on its magnitude.

5. Overall evaluation of coalition formation

In this section, we pull together results from the first and second stage in order to evaluate and compare the success in the M- and M + A-games overall.

5.1. General properties and results

Proposition 5 summarizes what we know about mitigation and adaptation strategies as well as payoffs for a generic coalition of size p and if a coalition of size $p-1$ is enlarged to p. This information will be helpful in comparing outcomes in the two games.

Proposition 5 (General properties of the overall game**).**

Let $n \geq p \geq 2$.

a) In both games, $Q^*(p) > Q^*(p-1)$ for all p, $2 \leq p \leq n$.
b) In both games, if $A \geq 0$, then $\Pi^*(p) > \Pi^*(p-1)$ for all p, $2 \leq p \leq n$ where Π is the total payoff, i.e. the sum of payoffs over all players.
c) Let $B_{Qx} > 0$:
i) $Q^{*M+A}(p) > Q^{*M}(p)$ for all p, $2 \leq p \leq n$.
ii) In the M + A-game, $x^*(p) > x^*(p-1)$ for all p, $2 \leq p \leq n$.
iii) For any coalition of size p, $2 \leq p \leq n$: $\Pi^{*M+A}_{i\in P}(p) > \Pi^{*M}_{i\in P}(p)$, $\Pi^{*M+A}_{j\notin P}(p) > \Pi^{*M}_{j\notin P}(p)$ and hence $\Pi^{*M+A}(p) > \Pi^{*M}(p)$.
d) Let $B_{Qx} < 0$:
i) $Q^{*M+A}(p) < Q^{*M}(p)$ for all p, $2 \leq p \leq n$.
ii) In the M + A-game, $x^*(p) < x^*(p-1)$ for all p, $2 \leq p \leq n$.
iii) For any coalition size p, $2 \leq p \leq n$, if the cost function is a strictly convex polynomial function and if $\Pi^{*M+A}_{j\notin P}(p) > \Pi^{*M}_{j\notin P}(p)$, then $\Pi^{*M+A}_{i\in P}(p) > \Pi^{*M}_{i\in P}(p)$.

Proof. See Appendix A.4 and the discussion below. ■

Result *a* highlights that total mitigation increases with the size of the coalition and hence achieves its maximum in the grand coalition, corresponding to the social optimum. That is, even if there is leakage, it will never be so strong as to compensate for the increase of mitigation through signatories, which goes along with an enlargement of the coalition. Moreover, despite adaptation is available in the M + A-game, with a growing coalition, total mitigation increases.

Result *b* shows that a corresponding property can be established for total payoffs if $A \geq 0$. Clearly, the very fact that we are dealing with an externality game implies that the largest total welfare is obtained in the grand coalition (also called cohesiveness), in which all externalities are internalized by definition. However, whether each expansion of a coalition increases the total payoff is a stronger property (also called full cohesiveness) but does not always hold. For instance, in negative externality games the negative effect on outsiders may be so strong that the total payoff may decrease with the size of the coalition for smaller or intermediate coalitions. But even in positive externality games, the expansion of coalitions may not lead to higher total payoffs if superadditivity fails and if this effect is pronounced enough. In contrast, if superadditivity holds, for which a sufficient condition is $A \geq 0$ as we know from Proposition 3, then any expansion of the coalition will increase total welfare.

The property of full cohesiveness is interesting for our analysis for at least two reasons. Firstly, it normatively motivates the search for large stable coalitions even if the grand coalition may not be stable. Secondly, it can facilitate comparison of the two games as will become apparent below.

Results *c* and *d* differ regarding the sign of the cross derivative B_{Qx}. Items *i* and *ii* are mirror image results whereas for $B_{Qx} > 0$ item *iii* allows for stronger results than for $B_{Qx} < 0$.

If mitigation and adaptation are substitutes, i.e. $B_{Qx} < 0$, mitigation levels will be higher in the M- than in the M + A-game for any generic coalition of size p. So even if stable coalitions are smaller in the M- than in the M + A-game, we cannot conclude that total mitigation will be higher in the M + A-game than in the M-game. Moreover, with increasing degrees of cooperation, adaptation levels will decrease and obtain its minimum in the grand coalition.

In contrast, if mitigation and adaptation are complements, i.e. $B_{Qx} > 0$, adaptation (and mitigation) will increase with the degree of cooperation and total mitigation will always be higher in the M + A- than in the M-game for the same degree of cooperation. Consequently, if stable coalitions are larger in the M + A- than in the M-game, so will be total mitigation levels.

In order to compare payoffs across the two games, we conduct the following exercise. Like the size of coalitions p, adaptation x can be treated as a (positive) parameter in our model such that we can write $q_{j\notin P}(x)$, $q_{i\in P}(x)$, $Q(x)$, $\Pi_{j\notin P}(x)$ and $\Pi_{i\in P}(x)$, noticing that by setting $x = 0$, all these variables and functions are those in the M-game. After some appropriate rearrangement of terms, which are spelled out in Appendix A.4, and showing that *sign* $\left(\frac{\partial q_{j\notin P}(p)}{\partial x(p)}\right) =$ *sign* (B_{Qx}), we find:

$$\frac{\partial \Pi_{j\notin P}}{\partial x(p)} = [B_x - D_x] + \left[B_Q\left(p^2\frac{C_{qq}(q_{j\notin P}(p))}{C_{qq}(q_{i\in P}(p))} + (n-p-1)\right)\frac{\partial q_{j\notin P}(p)}{\partial x(p)}\right] \tag{10}$$

$$\frac{\partial \Pi_{i\in P}}{\partial x(p)} = [B_x - D_x] + \left[B_Q\,(n-p)\,\frac{\partial q_{j\notin P}(p)}{\partial x(p)}\right]. \tag{11}$$

The first term in square brackets, $B_x - D_x$, which is the same for non-signatories and signatories, is positive at $x = 0$, decreases with x, becomes zero at $x^*(p)$ and will be negative for larger levels of x. The sign of the second term in square brackets, though different for non-signatories and signatories, depends only on the sign of $\frac{\partial q_{j\notin P}(p)}{\partial x(p)}$ which in turn depends only on the sign of B_{Qx} (because $B_Q\left(p^2\frac{C_{qq}(q_{j\notin P}(p))}{C_{qq}(q_{i\in P}(p))} + (n-p-1)\right)$ in Eq. (10) and B_Q in Eq. (11) are

positive). Let $x_k^{**}(p)$ denote the level where $\frac{\partial \pi_k}{\partial x(p)} = 0$ which is normally different from the equilibrium $x^*(p)$ as long as the grand coalition has not formed.

Now consider first the simple case of $B_{Qx} > 0$ and hence $\frac{\partial q_{j \notin P}(p)}{\partial x(p)} > 0$. Then, it is immediately clear that the second term in Eqs. (10) and (11) is positive (and zero for signatories at $n = p$). The first term $B_x - D_x$ is zero at $x^*(p)$. Hence, both derivatives in Eqs. (10) and (11) are positive if $p < n$. Hence, non-signatories and signatories are always better off in the M + A- than in the M-game. This is not very surprising as both strategies work in the same direction and reinforce their positive impact. More adaptation increases the benefits from mitigation (and vice versa) and hence having adaptation as an additional strategy at their disposal benefits all countries.

Probably more interesting is to consider the case of $B_{Qx} < 0$ and hence $\frac{\partial q_{j \notin P}(p)}{\partial x(p)} < 0$. Then the second term in brackets in Eqs. (10) and (11) is clearly negative. The first term is again zero at the equilibrium adaptation level $x^*(p)$. Thus, $\frac{\partial \pi_k}{\partial x(p)} < 0$ at $x^*(p)$ for signatories and non-signatories. In other words, $x^*(p)$ is always larger than optimal adaptation $x_k^{**}(p)$ as long as the grand coalition has not formed. Hence, $x^*(p)$ is located on the downward sloping part of the payoff function if expressed as a function of x as done above. Therefore, it is not straightforward to conclude whether players are better off or worse off in the M + A- than in the M-game as long as the grand coalition has not formed.

Nevertheless, some further observations are informative. Firstly, because $p^2 \frac{C_{qq}(q_{j \notin P}(p))}{C_{qq}(q_{i \in P}(p))} + (n - p - 1) > n - p$ (if the cost function is a strictly convex polynomial function; see proof in Appendix A.4), we have $\frac{\partial \Pi_{i \in P}}{\partial x(p)} > \frac{\partial \Pi_{j \notin P}}{\partial x(p)}$ (because $\frac{\partial q_{j \notin P}(p)}{\partial x(p)} < 0$). So if both derivatives in Eqs. (10) and (11) are positive at $x = 0$, and if we let $\bar{x}$ denote the level of adaptation which makes players indifferent between the M- and M + A-game, i.e. $\Pi_{j \notin P}(\bar{x}_{j \notin P}(p)) = \Pi_{j \notin P}(0)$ and $\Pi_{i \in P}(\bar{x}_{i \in P}(p)) = \Pi_{i \in P}(0)$, then we have $\bar{x}_{j \notin P}(p) < \bar{x}_{i \in P}(p)$. Consequently, if $x^*(p)$ is smaller than $\bar{x}_{j \notin P}(p)$, both groups of players are better off in the M + A-game, if $x^*(p)$ is larger than $\bar{x}_{i \in P}(p)$ both players are worse off and finally, if $x^*(p)$ lies between $\bar{x}_{j \notin P}(p)$ and $\bar{x}_{i \in P}(p)$ signatories are better off but non-signatories are worse off in the M + A- than in the M-game, which is result *d, iii* in Proposition 5.

It is important to recall that such comparisons are only valid for a given p. That is, a group of players could be better or worse off for some p but this could be different for another p. Clearly, if $p = n$, all players are signatories, $x^*(n) = x^{**}(n)$ and hence $\frac{\partial \Pi_{i \in P}(p)}{\partial x(p)} = 0$ in (11), and so all players are better off in the M + A- than in the M-game. Axiomatically, if there is no strategic interaction among players, having an additional strategy cannot make things worse, but will strictly improve in the context of an externality game.

It is worthwhile to pause for a second in order to understand the driving forces why players could be worse off in the M + A- than in the M-game provided mitigation and adaptation are substitutes as long as the grand coalition has not formed.

Firstly, on the "minus side", adaptation reduces the benefits of mitigation. This is because the marginal benefits are reduced but also because in equilibrium mitigation levels are lower in the M + A- than in the M-game. That is, the non-excludable benefits from mitigation are lower in the M + A- than in the M-game. In other words, the positive externalities across countries are lower.

Secondly, on the "plus side", adaptation offers a second strategy to address climate change. It directly increases benefits and reduces the aggregate cost. Thirdly, also on the "plus side", adaptation reduces the mitigation cost differential between signatories and non-signatories, $C(q_{i \in P}) - C(q_{j \notin P})$, as we show in Appendix A.4. For a given level Q, it is clear that from a cost-effectiveness point of view, non-signatories should mitigate more and signatories less until marginal costs equalize (which for symmetric players means the same mitigation level in our model). Through adaptation both types of players reduce their mitigation efforts, and hence the cost differential decreases due to the strict convexity of cost functions.

So overall, we have three effects of which the first goes in the opposite direction of the second and third effect and therefore predictions at a general level are difficult. We show below that the positive effects always dominate the negative effect for payoff function 1 but this is not always the case for payoff function 2.

5.2. Specific payoff functions

Comparing equilibrium payoffs of the overall game in the M- and M + A-games is analytically challenging. The comparison is less cumbersome for payoff function 1 because it is by construction simpler than payoff function 2. Results are summarized in Proposition 6.

Proposition 6 (Results of the overall game for payoff functions 1 and 2).

Payoff Function 1
$\Pi^{*M+A}(p^{*M+A}) > \Pi^{*M}(p^{*M})$ *irrespective of the sign of* B_{Qx} *and* $Q^{*M+A}(p^{*M+A}) > Q^{*M}(p^{*M})$ *if* $B_{Qx} > 0$.

Payoff Function 2
i) Let $B_{Qx} > 0$:
If $A^{M+A} \geq 0$, *then* $\Pi^{*M+A}(p^{*M+A}) > \Pi^{*M}(p^{*M})$ *and* $Q^{*M+A}(p^{*M+A}) > Q^{*M}(p^{*M})$.
If $A^{M+A} < 0$, *then* $\Pi^{*M+A}(p^{*M+A}) > \Pi^{*M}(p^{*M})$ *and* $Q^{*M+A}(p^{*M+A}) > Q^{*M}(p^{*M})$ *if* $p^{*M+A} \geq p^{*M}$.
ii) Let $B_{Qx} < 0$:
If $A^{M+A} \geq 0$, *then* $\Pi^{*M+A}(p^{*M+A}) > \Pi^{*M}(p^{*M})$ *if n is sufficiently large;* $Q^{*M+A}(p^{*M+A}) >< Q^{*M}(p^{*M})$.
If $A^{M+A} < 0$, *then* $\Pi^{*M+A}(p^{*M+A}) >< \Pi^{*M}(p^{*M})$ *and* $Q^{*M+A}(p^{*M+A}) >< Q^{*M}(p^{*M})$.

Proof. See Appendix A.5 and the discussion below. ■

For Payoff Function 1, we firstly know that $p^{*M+A} \geq p^{*M}$ (Proposition 4). Hence, $Q^{*M+A}(p^{*M+A}) > Q^{*M}(p^{*M})$ follows for $B_{Qx} > 0$ (using Proposition 5, result *c, i*). But it is also clear that for $B_{Qx} < 0$, the reverse is possible, for instance if $p^{*M+A} = p^{*M}$ (because Proposition 5, result *d, i*). Secondly, we know $A^M = 0$ and $A^{M+A} \geq 0$ (Proposition 4), implying full cohesiveness in both games (Proposition 5, result *b*). Therefore, $\Pi^{*M+A}(p^{*M+A}) > \Pi^{*M}(p^{*M})$ for $B_{Qx} > 0$ follows directly from Proposition 5, result *b* and *c, iii*. For $B_{Qx} < 0$, a payoff comparison is less obvious. However, in Appendix A.5 we show that for Payoff Function 1, signatories are always better off in the M + A- than in the M-game for every p, $1 \leq p \leq n$. For non-signatories, interestingly, we find that for sufficiently large p (though $p < n$) they could be worse off in the M + A- than in the M-game, depending on the values of the parameters. Examples are provided in Appendix A.5. However, we can show in Appendix A.5 that the total payoff will always be larger in M + A- than in the M-game for stable coalitions.

For Payoff Function 2, results are straightforward as long as $B_{Qx} > 0$. If $A^{M+A} \geq 0$, $p^{*M+A} > p^{*M}$ from Proposition 4 and hence Proposition 5 directly applies (for mitigation results *a* and *c, i* and for payoff results *b* and *c, iii*). For $A^{M+A} < 0$ (recall $A^M < 0$ always in the M-game), stable coalitions in both games are either the singleton coalition or a coalition of two players. If stable coalitions are of the same size in both games (i.e. $p^{*M+A} = p^{*M} = 1$ or $p^{*M+A} = p^{*M} = 2$), Proposition 5 gives immediately the result. If $p^{*M+A} = 2$ and $p^{*M} = 1$, conclusions for mitigation also follow from Proposition 5 *c, i*. For payoffs, as $A^{M+A} < 0$, we cannot assume full cohesiveness in general. However, as mentioned above in Section 4.1, if a coalition p is internally stable, the move from $p - 1$ to p is superadditive and therefore, because of positive externalities, this moves is also associated with an increase of total payoffs. Consequently,

$\Pi^{*M+A}(p^{*M+A} = 2) > \Pi^{*M}(p^{*M} = 1)$. But of course, we cannot rule out the possibility $p^{*M+A} = 1$ and $p^{*M} = 2$ in which case no general conclusions are possible.

Finally, less can be said for Payoff Function 2 and $B_{Qx} < 0$ at a general level. For $A^{M+A} \geq 0$, we know $p^{*M+A} > p^{*M}$. Not surprisingly, total mitigation levels in stable coalitions could be higher or lower in the M + A- than in the M-game, depending on parameter values. Regarding total payoffs, all our simulations which we have conducted (and on which report in our working paper) show $\Pi^{*M+A}(p^{*M+A}) > \Pi^{*M}(p^{*M})$. This is in line with our analytical proof on which we report in Appendix A.5, though this requires to assume that the number of players n is sufficiently large. For $A^{M+A} < 0$, and given the previous discussion, it is not surprising that there is no general ranking, neither regarding total mitigation nor total payoffs. Stable coalitions may be of equal or smaller size in the M + A-game than in the M-game and hence adaptation may not only lead to worse outcomes in terms of mitigation levels but also in terms of total payoffs. Appendix A.5 provides examples.

Taken together if mitigation and adaptation are complements, adaptation improves in terms of total mitigation and payoffs. If they are substitutes and if cross effects are strong enough (so that $A \geq 0$ in the M + A-game), adaptation also helps to increase total payoffs; mitigation levels can be larger but also be lower. However, if cross effects are not sufficiently strong (so that $A < 0$ in both games), no general conclusions are possible, but it is not unlikely that adaptation leads to worse outcomes, not only in terms of total mitigation levels but also in terms of total payoffs.

6. Conclusion

In a two-stage coalition formation model, we have analyzed how adaptation, as an additional strategy to mitigation, affects the prospects of international policy coordination to tackle climate change. Different from a pure mitigation game, adaptation may cause mitigation levels between different countries to be strategic complements. In those cases, there is no easy-riding but easy-matching. We showed that this may lead to larger stable coalitions with even the grand coalition being stable. Also, the associated global welfare may be higher with adaptation.

We showed that mitigation levels become strategic complements if the marginal benefits from mitigation (adaptation) are strongly affected by the level of adaptation (mitigation), i.e. the absolute value of the cross derivative of benefits with respect to mitigation and adaptation is sufficiently large. It is not important whether mitigation and adaptation are substitutes as commonly believed, or complements, but only that the rate of substitution or complementarity is large.

If mitigation and adaptation are complements, adaptation as an additional strategy to mitigation will always lead to larger stable agreements with larger global mitigation levels and global welfare. If they are substitutes, conclusions are less obvious. Regarding total mitigation, for a generic agreement (given coalition size), equilibrium mitigation will be lower with adaptation. Hence, adaptation must lead to a substantial increase in the size of stable agreements in order for equilibrium total mitigation to increase in the light of adaptation. Regarding global welfare, conclusions are not obvious because even though adaptation offers an additional strategy and hence reduces the total cost of addressing climate change, it also reduces the positive externalities across countries stemming from mitigation. For the specific payoff functions we considered, we could show that the overall effect works in favour of adaptation if the cross derivative is sufficiently large, implying upward sloping reaction functions in mitigation space. Finally, if the cross derivative is not large enough and hence reaction functions in mitigation space are downward sloping, adaptation may not only lead to worse outcomes in terms of global mitigation but also in terms of global welfare.

Our results hold if mitigation and adaptation are chosen simultaneously and also if mitigation is chosen first and then adaptation. Both versions are equivalent. We did not consider the version when adaptation is chosen first and then mitigation as in Zehaie (2009). A prediction how this may change our results is very difficult as Zehaie considers only two players and, more importantly, his framework does not allow for upward sloping reaction functions. What we can conclude from his alternative version is that if adaptation and mitigation are strategic substitutes, equilibrium mitigation for a given agreement size will be lower and adaptation higher than in our two versions. Though this may reduce the free-rider incentive (caused by ambitious mitigation targets) and may hence lead to larger stable coalitions, it may imply that less is achieved in terms of global welfare and global mitigation overall. In order to substantiate our conjecture, it would be interesting to analyze such a sequence in full detail in future research.

Our model made a couple of assumptions for analytical tractability and focus. For instance, we considered one of the most widespread coalition games and stability concepts (internal and external stability in a cartel formation game) but could have considered other concepts (Bloch, 2003; Finus and Rundshagen, 2009; Yi, 1997). Internal and external stability implies that after a player leaves the coalition, the remaining coalition members remain in the coalition. In the context of a positive externality game, this is the weakest possible punishment after a deviation and hence implies the most pessimistic assumption about stability. This appears to be a good benchmark because we could show that with adaptation larger coalitions can be stable, including the grand coalition.

What would certainly be interesting is to depart from the assumption of symmetric players in order to capture better the current discussion whether industrialized countries should support developing countries not only in their mitigation but also their adaptation efforts (Lazkano et al., 2016). Will support in adaptation buy more mitigation? In this context, one could assume that coalition members can pool their adaptation activities as a club, obtaining an additional benefit compared to non-signatories from the cost-effective production of adaptation. Essentially, this would require to model in kind-transfers apart from monetary transfers in a coalition formation model with heterogenous agents.

Appendix A

Equilibrium mitigation and adaptation levels depend on the size of a coalition p. Below, if not important, we drop this argument to shorten proofs.

A.1. Proof of Proposition 1

We follow Cornes and Hartley (2007) who have shown that replacement functions are a convenient and elegant tool to establish existence of a unique Nash equilibrium in aggregative games. From the first order conditions for mitigation, individual replacement function $q_{i \in P} = R_{i \in P}(Q)$ for signatories and by $q_{j \notin P} = R_{j \notin P}(Q)$ for non-signatories can be derived; the aggregate replacement function is given by $Q = R(Q) = \sum_{i \in N} R_i(Q)$. By definition, this identity defines the equilibrium. Graphically, the equilibrium is given by the intersection of the aggregate replacement function with the 45-degree line. If the aggregate replacement function is downward sloping over the entire domain, it must cross the 45-degree line, and does so only once. If the replacement function is upward sloping, a unique intersection does not follow automatically, though a sufficient condition

is that the aggregate replacement function has a slope less than one over the entire domain.

Consider the general formulation in the M + A-game. From $pB_Q(Q,x(Q)) = C_q(q_{i\in P})$, we have $q_{i\in P} = R_{i\in P}(Q) = C_q^{-1}(pB_Q(Q,x(Q)))$. $R_{i\in P}(Q)$ is continuous in Q and $q_{i\in P}$ is strictly positive if Q approaches zero because $\lim_{Q\to 0}B_Q > \lim_{q\to 0}C_q > 0$ from our General Assumptions. From the theorem of inverse functions, we have $\frac{dC_q^{-1}(q)}{dq} = \frac{1}{C_{qq}(q)}$ and hence the slopes of individual replacement functions of signatories are given by $R'_{i\in P}(Q) = \frac{d(C_q^{-1}(pB_Q(Q,x(Q))))}{dQ} = \frac{1}{C_{qq}(q_{i\in P})}\frac{d(pB_Q(Q,x(Q)))}{dQ} = \left(\frac{pA}{C_{qq}(q_{i\in P})}\right)$ because $\frac{d(B_Q(Q,x(Q)))}{dQ} = A$, and of non-signatories by $R'_{j\notin P}(Q) = \frac{1}{C_{qq}(q_{j\notin P})}A$, respectively. In more detail, $\frac{d(B_Q(Q,x(Q)))}{dQ} = B_{QQ} + B_{Qx}\frac{dx}{dQ}$ with $\frac{dx}{dQ} = \frac{B_{Qx}}{D_{xx}-B_{xx}}$ from $B_{xx}(Q,x)dx + B_{Qx}(Q,x)dQ - D_{xx}(x)dx = 0$ (total differentiation of the first-order conditions (Eq. (4)). For the aggregate replacement function, $Q = R(Q) = \sum_{i\in N}R_i(Q)$, accordingly, we derive $R'(Q) = A\left[\frac{p^2}{C_{qq}(q_{i\in P})} + \frac{(n-p)}{C_{qq}(q_{j\notin P})}\right]$. Finally, if Q^* is unique, $x^*(Q^*)$ is unique because $x = f_{i\in N}(Q)$ is continuous over the entire strategy space and $\lim_{x\to 0}B_x > \lim_{x\to 0}D_x > 0$ from our General Assumptions ensures an interior equilibrium. ***Q.E.D.***

A.2. Proof of claims in the text and Proposition 3

Superadditivity implies $p\Pi^*_{i\in P}(p) > [p-1]\,\Pi^*_{i\in P}(p-1) + \Pi^*_{j\notin P}(p-1)$.

1) *If a coalition of size $p \geq 2$ is internally stable, then the move from $p-1$ to p is superadditive.* The condition for superadditivity above can be rearranged such that we have: $\Pi^*_{i\in P}(p) + [p-1]\cdot(\Pi^*_{i\in P}(p) - \Pi^*_{i\in P}(p-1)) > \Pi^*_{j\notin P}(p-1)$. We note that internal stability implies $\Pi^*_{i\in P}(p) \geq \Pi^*_{j\notin P}(p-1)$. Moreover, $\Pi^*_{j\notin P}(p-1) > \Pi^*_{i\in P}(p-1)$ is generally true because for any p, we have $\Pi^*_{j\notin P}(p) = \Pi^*_{i\in P}(p) + \left[C(q^*_{i\in P}(p)) - C(q^*_{j\notin P}(p))\right]$ where the term in square brackets is positive because $q^*_{i\in P}(p) > q^*_{j\notin P}(p)$ and strictly convex cost functions, and hence $\Pi^*_{i\in P}(p) - \Pi^*_{i\in P}(p-1) > 0$ must hold and therefore also the condition for superadditivity. Note that $q^*_{i\in P}(p) > q^*_{j\notin P}(p)$ follows (in both games) from the first order conditions for mitigation (Eq. (3) in the M-game and Eq. (4) in the M + A-game), implying $pC_q(q^*_{j\notin P}(p)) = C_q(q^*_{i\in P}(p))$ and recalling $C_{qq} > 0$ from our General Assumptions. (Remark: We also notice that superadditivity implies $\Pi^*_{i\in P}(p) - \Pi^*_{i\in P}(p-1) > 0$, irrespective whether a coalition of size p is internally stable. This implies that if say two coalitions are stable, p_1^* and p_2^* with $p_1^* > p_2^*$, p_1^* Pareto-dominates p_2^* in superadditive and positive externality games. All players which are signatories (non-signatories) in p_1^* and p_2^* are better off by the property superadditivity (positive externality) and all players who are non-signatories in p_2^* but signatories in p_1^* are at least equally well off because $\Pi^*_{k\in P}(p_1^*) \geq \Pi^*_{k\notin P}(p_1^*-1) \geq \Pi^*_{k\notin P}(p_2^*)$ where the first inequality follows from internal stability and the second from positive externalities. We will make use of this Pareto-dominance property below).
2) *If the move from $p = 1$ to $p = 2$ is superadditive, a non-trivial stable coalition exists.* Let $p = 2$, in which case $\Pi^*_{i\in P}(p-1) = \Pi^*_{j\notin P}(p-1)$ and hence the condition for superadditivity stated above becomes $2\Pi^*_{i\in P}(p) \geq 2\Pi^*_{j\notin P}(p-1)$ or $\Pi^*_{i\in P}(p) \geq \Pi^*_{j\notin P}(p-1)$ which is the condition for internal stability. Hence, $p = 2$ is internally stable if the move from $p-1 = 1$ to $p = 2$ is superadditive. If $p = 2$ is externally stable, we are done. If not, then $p = 3$ must be internally stable. Repeating this argument means that, eventually, a coalition must be externally stable, noting that the grand coalition is externally stable by definition.
3) *A sufficient condition for superadditivity in both games is $A \geq 0$.* Consider the M + A-game and the condition for superadditivity as stated at the beginning above. Step 1: On the right-hand side of the inequality, the equilibrium values are $Q^*(p-1)$, $q^*_{i\in P}(p-1)$, $q^*_{j\notin P}(p-1)$ and $x^*(p-1)$ with $q^*_{i\in P}(p-1) > q^*_{j\notin P}(p-1)$. Now, we deduct ε from all signatories' mitigation levels and set one non-signatory j's mitigation level to exactly the same value, i.e. $\tilde{q}^*_{j\notin P}(p-1) = q^*_{i\in P}(p-1) - \epsilon$, keeping all other non-signatories mitigation level constant, choosing ε such that $Q^*(p-1)$ does not change. (Hence, $\varepsilon = \frac{\left(q^*_{i\in P}(p-1) - q^*_{j\notin P}(p-1)\right)}{p}$.) Hence, benefits do not change, but costs will drop because $pC(q^*_{i\in P}(p-1) - \epsilon) < [p-1]\,C(q^*_{i\in P}(p-1)) + C(q^*_{j\notin P}(p-1))$. We denote the payoff derived from the marginal change in step 1 for the p players by $\Pi^{\sup(1)}_{i\in P}(p)$ and hence can conclude $p\Pi^{\sup(1)}_{i\in P}(p) > (p-1)\Pi^*_{i\in P}(p-1) + \Pi^*_{j\notin P}(p-1)$. Step 2: If $A \geq 0$, for all other non-signatories $k \notin j$, $q^*_{k\notin P}(p-1) \leq q^*_{k\notin P}(p)$ as we show below under point 5 and because $\frac{\partial \Pi_i}{\partial q_k} > 0$, we have from step 2, $\Pi^{\sup(2)}_{i\in P} \geq \Pi^{\sup(1)}_{i\in P}$. Step 3: $\max_{q,x}\ p\Pi_{i\in P}(p) = p\Pi^*_{i\in P}(p) \geq p\Pi^{\sup(2)}_{i\in P}$. (Hence, moving from the right-hand side to the left-hand side of the SAD-condition the aggregate payoff of the enlarged coalition increases because total costs among the p players decrease (step 1), all outsiders increase their mitigation level (step 2) and the players in the enlarged coalition can freely choose mitigation and adaptation (step 3)). A slight modification of this proof applies to the M-game for $A = 0$.
4) *Both games are positive externality games.* We find generally $\frac{d\Pi^*_{j\notin P}}{dp} = B_Q\left[\frac{dQ^*}{dp}\left(1 - \frac{A}{C_{qq}(q_{j\notin P})}\right)\right] > 0$ because $\frac{dQ^*}{dp} > 0$ as we show below under point 5 and $\left(1 - \frac{A}{C_{qq}(q_{j\notin P})}\right) > 0$ from the condition for existence and uniqueness and the definition of A in Proposition 1.
5) $\frac{dQ^*}{dp} > 0$, *and* $\frac{dq_{j\notin P}}{dp} \geq (<)0$ *if* $A \geq (<)0$. Firstly, note that these derivatives treat p as a continuous variable, but if true, it is also true for discrete values. Secondly, we consider the general M + A-game, with obvious simplification for the M-game. Thirdly, in order to save on notation, we drop the asterisk and p as an argument in equilibrium mitigation and adaptation levels below. Moreover, we will omit the arguments in the benefit function because they are the same for signatories and non-signatories, but keep them for the mitigation cost function because $q_{i\in P} > q_{j\notin P}$ for every p, $1 < p < n$. We use three pieces of information in order to obtain the following results. i) Total differentiation of the first order conditions (Eq. (4)), $pB_Q(Q, x(Q)) = C_q(q_i)$, using $x(Q)$ and recalling from Appendix A.1 that $\frac{dB_Q(Q,\ x(Q))}{dQ} = A$, gives $dp\cdot B_Q + p\cdot A\cdot dQ = C_{qq}(q_{i\in P})\cdot dq_i$ for signatories with $p \geq 2$ and $A\cdot dQ = C_{qq}(q_{j\notin P})\cdot dq_{j\notin P}$ for non-signatories with $p = 1$. ii) Noting that the first order conditions (Eq. (4)) imply $p\cdot C_q(q_{j\notin P}) = C_q(q_{i\in P})$, total differentiation gives $dp\cdot C_q(q_{j\notin P}) + p\cdot C_{qq}(q_{j\notin P})\cdot dq_{j\notin P} = C_{qq}(q_{i\in P})\cdot dq_{i\in P}$. iii) $\frac{dQ}{dp} = q_{i\in P} + p\cdot\frac{dq_{i\in P}}{dp} - q_{j\notin P} + (n-p)\cdot\frac{dq_{j\notin P}}{dp}$.
Now, some basic manipulations lead to the following results. $\frac{dq_{j\notin P}}{dp} = \frac{A}{C_{qq}(q_{j\notin P})}\frac{dQ}{dp}$, implying because $C_{qq}(q_{j\notin P}) > 0$ and $\frac{dQ}{dp} > 0$ as we show below that the sign depends on the sign of A. $\frac{dq_{i\in P}}{dp} = \frac{B_Q}{C_{qq}(q_{i\in P})} + \frac{dQ}{dp}\frac{Ap}{C_{qq}(q_{i\in P})}$. Inserting $\frac{dq_{i\in P}}{dp}$ and $\frac{dq_{j\notin P}}{dp}$ in $\frac{dQ}{dp}$ as given in *iii* above and rearranging terms (noticing $B_Q = C_q(q_{j\notin P})$ from the first order condition), we have $\frac{dQ}{dp} = \frac{(q_{i\in P} - q_{j\notin P}) + p\frac{C_q(q_{j\notin P})}{C_{qq}(q_{i\in P})}}{1 - A\left[\frac{p^2}{C_{qq}(q_{i\in P})} + \frac{(n-p)}{C_{qq}(q_{j\notin P})}\right]} > 0$ because the nominator is positive due to $q_{i\in P} > q_{j\notin P}$ and the denominator due the condition of existence and uniqueness in Proposition 1. ***Q.E.D.***

A.3. Proof of Proposition 4

A.3.1. Preliminaries: payoff function 1

Let us assume first $B_{Qx} < 0$. In the M + A-game, we have: $B_Q = b - \Lambda x_i$ with $\Lambda = b\gamma + a\lambda$, $B_{QQ} = 0$, $B_{Qx} = -\Lambda < 0$, $B_x = a - \Lambda Q$, $B_{xx} = 0$, $C_q = cq_i$, $C_{qq} = c$, $D_x = dx_i$ and $D_{xx} = d$. Note that $A = B_{QQ} + \frac{(B_{Qx})^2}{D_{xx}-B_{xx}} = \frac{(-\Lambda)^2}{d} = \frac{\Lambda^2}{d} > 0$ and the condition for existence and uniqueness is most restrictive if $p = n$: $\frac{An^2}{C_{qq}} < 1 \iff cd - n^2\Lambda^2 > 0$. Equilibrium mitigation and adaptation levels are given by $q^*_{i\in P}(p) = \frac{p(db-\Lambda a)}{cd-\Lambda^2(n-p+p^2)}$, $q^*_{j\notin P}(p) = q^*_{i\in P}(p)/p$ and $x^*(p) = \frac{ca-b\Lambda(n-p+p^2)}{cd-\Lambda^2(n-p+p^2)}$ noting that $n^2 \geq n - p + p^2$ for $p \leq n$ with $n - p + p^2$ increasing in p.

We need to assume the following conditions to hold:

- C1: $1 - \gamma x_i > 0$ where x_i takes on the largest value in the Nash equilibrium;
- C2: $1 - \lambda Q > 0$ where Q takes on the largest value in the social optimum;
- C3: $B_Q = b - \Lambda x_i > 0$ where x_i takes on the largest value in the Nash equilibrium;
- C4: $B_x = a - \Lambda Q > 0$ where Q takes on the largest value in the social optimum;
- C5: $cd - n^2\Lambda^2 > 0$;
- C6: $db - \Lambda a > 0$;
- C7: $ca - n^2 b\Lambda > 0$

where C1 and C2 are required for the payoff function to make sense, C3 and C4 are required to be in line with the General Assumptions, C5 is the sufficient condition for existence and uniqueness of a second stage equilibrium and C6 and C7 are required to have an interior equilibrium for mitigation and adaptation, respectively, for every p, $1 \leq p \leq n$. We note that C1 and C2 are redundant because of C3 and C4 respectively; inserting the maximum values in C3 and C4, it is apparent that these two conditions are captured by C6 and C7, respectively, and hence can be dropped. Finally, C6 and C7 imply C5; hence, we are left with two conditions, namely C6 and C7. Now in order to capture $B_{Qx} > 0$, we simply set γ to $-\gamma$, λ to $-\lambda$ and hence Λ is $-\Lambda$. Hence, all conditions above automatically hold and term A does not change.

In the M-game, we find $q^*_{i\in P}(p) = \frac{pb}{c}$ and $q^*_{j\notin P}(p) = q^*_{i\in P}(p)/p$ where no conditions need to be imposed.

A.3.2. Preliminaries payoff function 2

Let us assume first $B_{Qx} < 0$. In the M + A-game, we have: $B_Q = b - gQ - fx_i$, $B_{QQ} = -g < 0$, $B_{Qx} = -f < 0$, $B_x = a - fQ$, $B_{xx} = 0$, $C_q = cq_i$, $C_{qq} = c$, and $D_x = dx_i$ and $D_{xx} = d$. Note that $A = B_{QQ} + \frac{(B_{Qx})^2}{D_{xx}-B_{xx}} = -g + \frac{(-f)^2}{d} = \frac{f^2-gd}{d}$ with $A < 0$ if $f^2 - gd < 0$ and $A \geq 0$ if $f^2 - gd \geq 0$; the condition for existence and uniqueness is most restrictive if $p = n$: $\frac{An^2}{C_{qq}} < 1 \iff cd - n^2(f^2 - gd) > 0$. Moreover, $q^*_{i\in P}(p) = \frac{p(bd-af)}{cd-(n-p+p^2)(f^2-gd)}$, $q^*_{j\notin P}(p) = q^*_{i\in P}(p)/p$ and $x^*(p) = \frac{ca-(n-p+p^2)(bf-ga)}{cd-(n-p+p^2)(f^2-gd)}$, noting that $n^2 \geq n - p + p^2$ for $p \leq n$ with $n - p + p^2$ increasing in p.

We need to assume the following conditions to hold:

- C1: $b - gQ - fx_i > 0$ where Q (resp. x_i) takes on the largest value in the social optimum (resp. Nash equilibrium);
- C2: $a - fQ > 0$ where Q takes on the largest value in the social optimum;
- C3: $cd - n^2(f^2 - gd) > 0$;
- C4: $bd - af > 0$;
- C5: $ca - n^2(bf - ga) > 0$

where C1 and C2 are required to be in line with the General Assumptions, C3 is the sufficient condition for existence and uniqueness of an equilibrium, C4 and C5 are required to have an interior equilibrium for every p, $1 \leq p \leq n$. Inserting the maximum values in C1 and C2, it turns out that C4 captures C1, C5 captures C2 and hence C1 and C2 can be dropped. So we are left with three conditions C3, C4 and C5. In order to capture $B_{Qx} > 0$, we simply set f to $-f$, noticing that this does not affect term A and that conditions C4 and C5 become redundant.

In the M-game, we have $q^*_{i\in P}(p) = \frac{pb}{p^2g+c}$ and $q^*_{j\notin P}(p) = q^*_{i\in P}(p)/p$ where no conditions need to be imposed.

A.3.3. Stability of payoff functions 1 and 2

Consider first the M-game. For payoff function 1, $p^* = 2$ and $p^* = 3$ is well-known in the literature where the latter Pareto-dominates the former equilibrium. For payoff function 2, internal stability in the M-game is given by:

$$IS(p) = -\frac{b^2c(p-1)\Omega(p)}{2(c+gp^2+gn-gp)^2(c+gp^2-3gp+2g+gn)^2}$$

where the denominator is clearly positive. $IS(p) \geq 0$ iff $\Omega(p) \leq 0$ where

$$\begin{aligned}\Omega(p) = {} & p^5g^2 - 5g^2p^4 + 2p^3cg + 2p^3ng^2 + 7g^2p^3 - 8cgp^2 - 4g^2p^2n \\ & -3g^2p^2 + n^2g^2p - 2g^2pn + 6cgp + 2npcg + pc^2 - 2cgn - 3c^2 \\ & -4gc + g^2n^2. \end{aligned} \tag{7}$$

Now we have:

$$\begin{aligned}\frac{\partial\Omega(p)}{\partial p} = {} & 5g^2p^3(p-4) + cgp(6p-16) + g^2pn(6p-8) + g^2p(21p-6) \\ & + g^2n(n-2) + (6gc + 2cgn + c^2)\end{aligned}$$

which is clearly positive for $p \geq 4$. Inserting $p = 3$, $p = 2$ and $p = 1$ will also confirm that this term is positive, i.e. $\frac{\partial\Omega}{\partial p} > 0$ for all p, $1 \leq p \leq n$. Hence, we check Ω for $p = 3$ and find:

$$12ng^2 - 4cg + 4g^2n^2 + 4cgn > 0$$

for $n \geq 3$ and hence $IS < 0$ for $p = 3$ and hence $p = 2$ is externally stable. For $p = 2$, it is easily checked that $\Omega \leq 0$ and hence $IS \geq 0$ is possible, depending on the parameter values. If $IS(2) > 0$, $p = 1$ is externally unstable and $p^* = 2$. If $IS(2) = 0$, $p^* = 2$ Pareto-dominates $p^* = 1$. Finally, if $IS(2) < 0$, $p^* = 1$.

Consider now stability in the M + A-game. Some simple though cumbersome manipulations gives

$$IS(p) = -\frac{(bd-af)^2cd^2(p-1)\Gamma(p)}{2Z}$$

with

$$\begin{aligned} Z : &= (cd + dgp^2 - p^2f^2 - nf^2 + ndg + pf^2 - pdg)^2 \\ & (cd + dgp^2 - 3pdg + 2dg - p^2f^2 + 3pf^2 - 2f^2 - nf^2 + ndg)^2 \\ \Gamma(p) : &= A^2\Psi_1(p) - Ac\Psi_2(p) + c^2(p-3) \\ \Psi_1(p) : &= p^5 + 2p^3n + pn^2 + n^2 + 7p^3 - 5p^4 - 3p^2 - 4p^2n - 2pn \\ \Psi_2(p) : &= 6p + 2p^3 + 2np - 8p^2 - 2n - 4 \end{aligned}$$

for payoff function 2. For payoff function 1, the same is true if we set $g = 0$ and replace f by Λ as defined above. Term A has been derived above for both payoff functions. Because $Z > 0$, for $p > 1$, $IS(p) \geq 0$ if $\Gamma(p) \leq 0$ and $IS(p) < 0$ if $\Gamma(p) > 0$. One can show that $\Psi_1(p) > 0$ and $\Psi_2(p) > 0$ for all p, $1 \leq p \leq n$ and $n \geq 3$. Hence, it is

immediately clear that if $A < 0, \Gamma(p) > 0$ for $p \geq 3$ and hence $p^* < 3$. It is also immediately clear that if $A = 0$, i) $\Gamma(3) = 0$ and hence $IS(3) = 0$, ii) $\Gamma(2) < 0$ and hence $IS(2) > 0$ and iii) $\Gamma(p) > 0$ and $IS(p) < 0$ for $p > 3$. Because of Pareto-dominance (see the remark in Appendix A.2 under point 1), we have $p^* = 3$ for $A = 0$. So now we let $A > 0$ and $p = 3$. Then $\Gamma \leq 0$ if $A\frac{\Psi_1(3)}{\Psi_2(3)} \leq c$. This is the case. From the condition of existence and uniqueness (as stated above), we have $An^2 < c$ and $\frac{\Psi_1(3)}{\Psi_2(3)} \leq n^2$, or $\frac{4n^2+12n}{4n-4} \leq n^2$ or $0 \leq n^2 - 2n - 3$, which is true for $n \geq 3$. So $p^* = 3$ is always an equilibrium. There can be further larger equilibria, including the grand coalition, depending on the value of A. If they exist, the largest Pareto-dominates the smaller equilibria by the remark in Appendix A.2 under point 1. (Remark: Solving $\Gamma(p) = 0$ for $A(p)$ gives two roots, $A_1(p)$ and $A_2(p)$, which both exist and if $A_1(p) \leq A(p) \leq A_2(p)$, then $\Gamma(p) \leq 0$ and $IS(p) \geq 0$. Typically, for a given $A > 0$, there is at most one other equilibrium with $p^* > 3$ apart from $p^* = 3$ with the larger equilibrium Pareto-dominating the smaller equilibrium.) ***Q.E.D.***

A.4. *Proof of Proposition 5*

a) $\frac{dQ}{dp} > 0$. See Appendix A.3, point 5. b) $\frac{d\Pi}{dp} > 0$ *if* $A \geq 0$. From Proposition 3, we know that both games are positive externality games and that $A \geq 0$ implies that both games are superadditive games, which together implies full cohesiveness. c) and d) part *i.* $Q^{*M+A}(p) > (<) Q^{*M}(p)$ *for all* p, $2 \leq p \leq n$ *if* $B_{Qx} > (<) 0$. One shows that the contradiction $Q^{M+A*}(p) < Q^{M*}(p)$ in the case of $B_{Qx} > 0$ and $Q^{M+A*}(p) > Q^{M*}(p)$ in the case of $B_{Qx} < 0$ is false by using the first order conditions for mitigation (Eq. (3) in the M-game and Eq. (4) in the M + A-game), and $B_{QQ} \leq 0$ and $C_{qq} > 0$ from the General Assumptions, noting that $x > 0$ in an interior equilibrium. c) and d) part *ii.* $\frac{dx}{dp} > (<) 0$ *if* $B_{Qx} > (<) 0$. In the spirit of Appendix A.2, in particular point 5, we can derive $\frac{dx}{dp} = \frac{B_{Qx}}{D_{xx}-B_{xx}}\frac{dQ}{dp}$ where the sign depends on the sign of B_{Qx} because $\frac{dQ}{dp} > 0$. c) and d) part *iii.* All arguments have been developed in the text, except showing that $p^2\frac{C_{qq}(q_{j\notin P})}{C_{qq}(q_{i\in P})} + (n-p-1) > n-p$ holds which implies $p^2\frac{C_{qq}(q_{j\notin P})}{C_{qq}(q_{i\in P})} > 1$. We know from above $C_q(q_{i\in P}) = pC_q(q_{j\notin P})$. Consider the cost function $C = \frac{c}{\epsilon}q^{\varepsilon}$ with $C_q = cq^{\varepsilon-1}$ and $C_{qq} = (\varepsilon-1)cq^{\varepsilon-2}$ which gives for $p^2\frac{C_{qq}(q_{j\notin P})}{C_{qq}(q_{i\in P})} = p^{\frac{\epsilon}{\epsilon-1}}$ which is larger than 1 for any $\varepsilon \geq 1$ in line with the assumption of strictly convex cost functions.

Eqs. (10) and (11) in the text are derived as follows. From the first order conditions of signatories $pB_Q(Q,x) = C_q(q_{i\in P})$ and of non-signatories $B_Q(Q,x) = C_q(q_{j\notin P})$, $C_q(q_{i\in P}) = pC_q(q_{j\notin P})$, the existence of second derivatives different from zero, and the theorem of implicit functions, we have

$$\frac{\partial q_{j\notin P}(p)}{\partial x(p)} = \frac{-B_{Qx}}{B_{QQ}\left[p^2\frac{C_{qq}(q_{j\notin P}(p))}{C_{qq}(q_{i\in P}(p))} + n - p\right] - C_{qq}(q_{j\notin P}(p))}$$

$$\frac{\partial q_{i\in P}(p)}{\partial x(p)} = p\frac{C_{qq}(q_{j\notin P}(p))}{C_{qq}(q_{i\in P}(p))}\frac{\partial q_{j\notin P}(p)}{\partial x(p)}.$$

Moreover, we have:

$$\begin{aligned}\frac{\partial Q(p)}{\partial x(p)} &= p\frac{\partial q_{i\in P}(p)}{\partial x(p)} + (n-p)\frac{\partial q_{j\notin P}(p)}{\partial x(p)} \\ &= \left[p^2\frac{C_{qq}(q_{j\notin P}(p))}{C_{qq}(q_{i\in P}(p))} + n - p\right]\frac{\partial q_{j\notin P}(p)}{\partial x(p)}.\end{aligned}$$

Finally, we have

$$\frac{\partial \Pi_{j\notin P}(p)}{\partial x(p)} = B_x - D_x + B_Q\frac{\partial Q(p)}{\partial x(p)} - C_q(q_{j\notin P}(p))\frac{\partial q_{j\notin P}(p)}{\partial x(p)}$$

$$\frac{\partial \Pi_{i\in P}(p)}{\partial x(p)} = B_x - D_x + B_Q\frac{\partial Q(p)}{\partial x(p)} - C_q(q_{i\in P}(p))\frac{\partial q_{i\in P}(p)}{\partial x(p)}$$

which, using the information above, gives Eqs. (10) and (11) in the text.

We also claim in the text that cost differences in terms of mitigation between signatories and non-signatories decrease with adaptation if $B_{Qx} < 0$. $\frac{\partial[C(q_{i\in P})-C(q_{j\notin P})]}{\partial x} = C_q(q_{i\in P})\frac{\partial q_{i\in P}}{\partial x} - C_q(q_{j\notin P})\frac{\partial q_{j\notin P}}{\partial x}$; noting $\frac{\partial q_{i\in P}}{\partial x} = p\frac{Cqq(q_{j\notin P})}{Cqq(q_{i\in P})}\frac{\partial q_{j\notin P}}{\partial x}$ as derived above, and $C_q(q_{i\in P}) = pC_q(q_{j\notin P})$, we have $\frac{\partial[C(q_{i\in P})-C(q_{j\notin P})]}{\partial x} = \left[p^2\frac{Cqq(q_{j\notin P})}{Cqq(q_{i\in P})} - 1\right]C_q(q_{j\notin P})\frac{\partial q_{j\notin P}}{\partial x}$. The term in square brackets is positive as shown above and $\frac{\partial q_{j\notin P}}{\partial x} < 0$ if $B_{Qx} < 0$. ***Q.E.D.***

A.5. *Proof of Proposition 6*

Consider payoff function 1 and $B_{Qx} < 0$ with $q_{j\notin P}(x)$, $q_{i\in P}(x)$, $Q(x)$, $\Pi_{j\notin P}(x)$ and $\Pi_{i\in P}(x)$. Solving the first order conditions of adaptation, gives $q_{i\in P}(x) = \frac{p(b-\Lambda x)}{c}$, $q_{j\notin P}(x) = q_{i\in P}(x)/p$ and $Q(x) = pq_{i\in P}(x) + (n-p)q_{j\notin P}(x)$. Inserting this into payoff function Eq. (7a) in the text gives after some rearrangement

$$\begin{aligned}\Pi_{i\in P}(x) &= -ux^2 + vx + w \\ w &:= \frac{b^2}{2c}(2\tilde{n} - p^2) \\ v &:= \frac{1}{c}(ac - b\Lambda(2\tilde{n} - p^2)) \\ u &:= \frac{1}{2c}(cd - \Lambda^2(2\tilde{n} - p^2))\end{aligned}$$

where $\tilde{n} = n - p + p^2$. All parameters, w, v and u are positive (u by condition C5 and v by condition C7 in Appendix A.3, noticing that $2\tilde{n} - p^2 \leq n^2$. Thus, $\Pi_{i\in P}(x)$ is strictly concave in x and obtains its maximum at $x_i^{**} = \frac{v}{2u} > 0$. Moreover, $\Pi_{i\in P}^{M+A} = x^*(v - ux^*) + w > w = \Pi_{i\in P}^{M} \iff x^* < \frac{v}{u} = \bar{x}_i$. It is easy to check that $x^* > x_i^{**}$ (where x^*, the adaptation level at the equilibrium is given in Appendix A.3) and slightly more involved to show that $x^* < \frac{v}{u} = \bar{x}_i$ by using condition C5 and condition C7 in Appendix A.3. Hence, $\Pi_{i\in P}^{M+A} > \Pi_{i\in P}^{M}$ for every p. For non-signatories we find:

$$\begin{aligned}\Pi_{j\notin P}(x) &= -ux^2 + vx + w \\ w &:= \frac{b^2}{2c}(2\tilde{n} - 1) \\ v &:= \frac{1}{c}(ac - b\Lambda(2\tilde{n} - 1)) \\ u &:= \frac{1}{2c}(cd - \Lambda^2(2\tilde{n} - 1))\end{aligned}$$

noticing that $w > 0$. Moreover, for $p = 1$, we have $\Pi_{j\notin P}^{M+A}(x) = \Pi_{i\in P}^{M+A}(x)$ and hence $\Pi_{j\notin P}^{M+A} > \Pi_{j\notin P}^{M}$ due to what we know from above. That is for non-signatories, $v > 0$ and $u > 0$ for $p = 1$. However, for larger values of p , v and u are not necessarily positive, despite conditions C5 and C7 are assumed to hold. (This is because $2\tilde{n} - 1 > n^2$ is possible.) This depends on the specific parameter values. For instance, it is possible that $v > 0$ and $u < 0$, and hence $x_j^{**} = \frac{v}{2u} < 0$ is a minimum, $\Pi_{j\notin P}(x)$ is a convex function and hence for any $x > 0$, $\Pi_{j\notin P}^{M+A} > \Pi_{j\notin P}^{M}$. However, it is also possible that $v < 0$ and $u > 0$, and hence $x_j^{**} = \frac{v}{2u} < 0$ is a maximum, $\Pi_{j\notin P}(x)$ is a concave function and hence for any $x > 0$, $\Pi_{j\notin P}^{M+A} < \Pi_{j\notin P}^{M}$. Examples of the last constellations include for instance the following parameter values: $n = 10, a = 300, \gamma = \lambda = 0.1, b = 300$. Now if i) $c = 5350$, $d = 3330$ and $p = n - 1 = 9$, ii) $c = 4320$, $d = 3303$ and $p = n - 1 = 8$ and iii) $c = 3390$, $d = 3309$ and $p = n - 1 = 7$, then $\Pi_{j\notin P}^{M+A} < \Pi_{j\notin P}^{M}$.

Nevertheless, we can show for stable coalitions $\Pi^{*M+A}(p^{*M+A}) > \Pi^{*M}(p^{*M})$, recalling that $p^{*M+A} \geq p^{*M} = 3$. We do so by showing that

for $p = 3$, $\Pi^{M+A}_{j \notin P}(3) > \Pi^{M}_{j \notin P}(3)$ and because $\Pi^{M+A}_{i \in P} > \Pi^{M}_{i \in P}$ for all p anyway, we have $\Pi^{*M+A}(3) > \Pi^{*M}(3)$ and hence due to cohesiveness also $\Pi^{*M+A}(p^{*M+A}) > \Pi^{*M}(p^{*M})$ if $p^{*M+A} > p^{*M} = 3$. We can exclude $p = 3$ and $n = 3$ from the analysis because in the grand coalition there are no non-signatories and $\Pi^{*M+A}(n) > \Pi^{*M}(n)$ is always true. So we assume $p = 3$ and $n > 3$ in which case $v > 0$ and $u > 0$ for non-signatories because $2\tilde{n} - 1 < n^2$ and due to conditions C5 and C7. Moreover, one can show that $x^* < \frac{v}{u}$ for non-signatories under these conditions which completes the proof.

Consider payoff function 2 and $B_{Qx} < 0$. To show that $\Pi^{*M+A}(p^{*M+A}) < \Pi^{*M}(p^{*M})$ is possible if $A^{M+A} < 0$, consider the following example: $n = 10$, $a = 10{,}000$, $b = 4000$, $c = 3000$, $d = 4000$, $f = 600$ and $g = 220$ implying $A = -130$, $p^{*M+A} = 1$, $p^{*M} = 2$ and $\Pi^{*M+A}(1) - \Pi^{*M}(2) = -5372.45$; none of the conditions C3, C4 and C5 is violated. It easy to find other examples for which $\Pi^{*M+A}(p^{*M+A}) < \Pi^{*M}(p^{*M})$ is true.

Finally, the claim that if $A^{M+A} \geq 0$, then $\Pi^{*M+A}(p^{*M+A}) > \Pi^{*M}(p^{*M})$ if n is sufficiently large, uses the information (as proved in the text) that if $\Pi^{M+A}_{j \notin P}(p) > \Pi^{M}_{j \notin P}(p)$, then $\Pi^{M+A}_{i \in P}(p) > \Pi^{M}_{i \in P}(p)$. Moreover, it uses the information that because $p^{*M+A} > p^{*M} = 2$ if $A^{M+A} \geq 0$, then we can conclude $\Pi^{M+A}(p^{*M+A}) > \Pi^{M+A}(2)$ due to full cohesiveness. Thus, it remains to be shown that $\Pi^{M+A}_{j \notin P}(2) > \Pi^{M}_{j \notin P}(2)$ for which we use C3 to C5 and assume n sufficiently large. This last step of the proof is extremely lengthy and hence available upon request. ***Q.E.D.***

References

d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J., Weymark, J.A., 1983. On the stability of collusive price leadership. Can. J. Econ. 16 (1), 17–25.

Barrett, S., 1994. Self-enforcing international environmental agreements. Oxf. Econ. Pap. 46, 878–894.

Barrett, S., 2008. Dikes vs. Windmills: Climate Treaties and Adaptation. Johns Hopkins University, School of Advanced International Studies.

Battaglini, M., Harstad, B., 2016. Participation and duration of environmental agreements. J. Polit. Econ. 124 (1), 160–204.

Baumol, W.J., Bradford, D.F., 1972. Detrimental externalities and non-convexity of the production set. Economica 39, 160–176.

Benchekroun, H., Marrouch, W., Chaudhuri, A.R., 2017. Adaptation Technology and Free-riding Incentives in International Environmental Agreements. In: Kayalca, O., Cagatay, S., Mihci, H. (Eds.), Economics of International Environmental Agreements: A Critical Approach. Routledge, UK, pp. 204–228. chapter 11.

Bergstrom, T., Blume, L., Varian, H., 1986. On the private provision of public goods. J. Public Econ. 29 (1), 25–49.

Bloch, F., 2003. Non-cooperative Models of Coalition Formation in Games with Spillovers. In: Carraro, C. (Ed.), The Endogenous Formation of Economic Coalitions. Edward Elgar, Cheltenham, UK, pp. 35–79. ch. 2.

Buob, S., Stephan, G., 2011. To mitigate or to adapt: how to confront global climate change. Eur. J. Polit. Econ. 27, 1–16.

Carraro, C., Siniscalco, D., 1993. Strategies for the international protection of the environment. J. Public Econ. 52 (3), 309–328.

Cornes, R., Sandler, T., 1984. Easy riders, joint production, and public goods. Econ. J. 94 (375), 580–598.

Cornes, R., Sandler, T., 1986. The Theory of Externalities, Public Goods and Club Goods. Cambrige University Press.

Cornes, R.C., Hartley, R., 2007. Aggregative public good games. J. Public Econ. Theory 9 (2), 201–219.

Deneckere, R., Davidson, C., 1985. Incentives to form coalitions with Bertrand competition. RAND J. Econ. 16 (4), 473–486.

Donsimoni, M.-P., Economides, N.S., Polemarchakis, H.M., 1986. Stable cartels. Int. Econ. Rev. 27 (2), 317–327.

Ebert, U., Welsch, H., 2011. Optimal response functions in global pollution problems can be upward-sloping: accounting for adaptation. Environ. Econ. Policy Stud. 13 (2), 129–138.

Ebert, U., Welsch, H., 2012. Adaptation and mitigation in global pollution problems: economic impacts of productivity, sensitivity, and adaptive capacity. Environ. Resour. Econ. 52, 49–64.

Eisenack, K., Kähler, L., 2016. Adaptation to climate change can support unilateral emission reductions. Oxf. Econ. Pap. 68 (1), 258–278.

El-Sayed, A., Rubio, S., 2014. Sharing R&D investments in cleaner technologies to mitigate climate change. Resour. Energy Econ. 38, 168–180.

Finus, M., Caparrós, A., 2015. Game Theory and International Environmental Cooperation: Essential Readings. Edward Elgar, Cheltenham.

Finus, M., Rundshagen, B., 2009. Membership rules and stability of coalition structures in positive externality games. Soc. Choice Welf. 32, 389–406.

Fraser, C.D., 1992. The uniqueness of Nash equilibrium in the private provision of public goods: an alternative proof. J. Public Econ. 49, 389–390.

Haeringer, G., 2004. Equilibrium binding agreements: a comment. J. Econ. Theory 117 (1), 140–143.

Harstad, B., 2012. Climate contracts: a game of emissions, investments, negotiations, and renegotiations. Rev. Econ. Stud. 79 (4), 1527–1557.

Kolstad, C.D., 2007. Systematic uncertainty in self-enforcing international environmental agreements. J. Environ. Econ. Manag. 53 (1), 68–79.

Laffont, J.-J., 1976. Decentralization with externalities. Eur. Econ. Rev. 7, 359–375.

Lazkano, I., Marrouch, W., Nkuiya, B., 2016. Adaptation to climate change: how does heterogeneity in adaptation costs affect climate coalitions? Environ. Dev. Econ. 21, 812–838.

Poyago-Theotoky, J., 1995. Equilibrium and optimal size of a research joint venture in an oligopoly with spillovers. J. Ind. Econ. 43 (2), 209–226.

Rubio, S.J., Ulph, A., 2006. Self-enforcing international environmental agreements revisited. Oxf. Econ. Pap. 58 (2), 233–263.

Starrett, D.A., 1972. Fundamental nonconvexities in the theory of externalities. J. Econ. Theory 4 (2), 180–199.

Vives, X., 2005. Complementarities and games: new developments. J. Econ. Lit. 43 (2), 437–479.

Weikard, H.-P., 2009. Cartel stability under an optimal sharing rule. Manch. Sch. 77 (5), 575–593.

Yi, S.-S., 1997. Stable coalition structures with externalities. Games Econ. Behav. 20 (2), 201–237.

Zehaie, F., 2009. The timing and strategic role of self-protection. Environ. Resour. Econ. 44 (3), 337–350.